Low scaling of a life history variable: Analysing eutherian gestation periods with and without phylogeny-informed statistics

Clauss, Marcus ; Dittmann, Marie T ; Müller, Dennis W H ; Zerbe, Philipp ; Codron, Daryl

Abstract: Traditionally, biological times (gestation period, longevity) are proposed to scale to body mass M as M0.25. Although phylogeny-informed statistics have become widespread, it is still sometimes assumed that in datasets comprising a very large number of species, analyses that do not and that do account for phylogeny will yield similar results. Here we show, in a large dataset on gestation period length in eutherian mammals (1214 species from 20 orders), that the allometric scaling exponent is about twice as high using conventional statistics (Ordinary Least Squares OLS, M0.18-0.20) as when using Phylogenetic Generalized Least Squares (PGLS, M0.07-0.10), indicating that among closely related taxa, the scaling of gestation is much lower than generally assumed. This matches the well-known absence of scaling among different-sized breeds of domestic animal species, and indicates that changes in M must be more related to changes in development speed rather than development time among closely related species, which has implications for interpreting life history-consequences of insular dwarfism and gigantism. Only when allowing just one species per order (simulated in 100 randomized datasets of n=20 species across 20 orders) is 0.25 included in the scaling exponent confidence interval in both OLS and PGLS. Differences in scaling at different taxonomic levels in comparative datasets may indicate evolutionary trends where successive taxonomic groups compete by fundamental variation in organismal design not directly linked to changes in M. Allometries then do not necessarily represent universal scaling rules, but snapshots of evolutionary time that depend on diversification and extinction events before the picture was taken. It is either by analysing subsets separately, or by using PGLS in large datasets, that the underlying relationships with M can then be unveiled.

DOI: https://doi.org/10.1016/j.mambio.2013.01.002

Mammalian Biology

Low scaling of a life history variable: analysing eutherian gestation periods with and without phylogeny-informed statistics

Marcus Clauss[1*], Marie T. Dittmann[2], Dennis W. H. Müller[1,3], Philipp Zerbe[1,4], Daryl Codron[1,5]

[1]Clinic for Zoo Animals, Exotic Pets and Wildlife, Vetsuisse Faculty, University of Zurich, Winterthurerstrasse 260, 8057 Zurich, Switzerland

[2]Institute of Plant, Animal and Agroecosystem Sciences, Swiss Federal Institute of Technology, Universitätsstrasse 2, 8092 Zurich, Switzerland

[3]National Park 'Bavarian Forest', Freyungerstr. 2, 94481 Grafenau, Germany

[4]Section for Small Animal Reproduction, Clinic for Animal Reproduction, Vetsuisse Faculty, University of Zurich, Winterthurerstrasse 260, 8057 Zurich, Switzerland

[5]Florisbad Quaternary Research, National Museum, Bloemfontein, 9300, RSA

Running head: Eutherian gestation scaling

Abstract

Traditionally, biological times (gestation period, longevity) are proposed to scale to body mass M as $M^{0.25}$. Although phylogeny-informed statistics have become widespread, it is still sometimes assumed that in datasets comprising a very large number of species, analyses that do not and that do account for phylogeny will yield similar results. Here we show, in a large dataset on gestation period length in eutherian mammals (1214 species from 20 orders), that the allometric scaling exponent is about twice as high using conventional statistics (Ordinary Least Squares OLS, $M^{0.18\text{-}0.20}$) as when using Phylogenetic Generalized Least Squares (PGLS, $M^{0.07\text{-}0.10}$), indicating that among closely related taxa, the scaling of gestation is much lower than generally assumed. This matches the well-known absence of scaling among different-sized breeds of domestic animal species, and indicates that changes in M must be more related to changes in development speed rather than development time among closely related species, which has implications for interpreting life history-consequences of insular dwarfism and gigantism. Only when allowing just one species per order (simulated in 100 randomized datasets of n=20 species across 20 orders) is 0.25 included in the scaling exponent confidence interval in both OLS and PGLS. Differences in scaling at different taxonomic levels in comparative datasets may indicate evolutionary trends where successive taxonomic groups compete by fundamental variation in organismal design not directly linked to changes in M. Allometries then do not necessarily represent universal scaling rules, but snapshots of evolutionary time that depend on diversification and extinction events before the picture was taken. It is either by analysing subsets separately, or by using PGLS in large datasets, that the underlying relationships with M can then be unveiled.

Key words: Metabolic Theory of Ecology; Reproduction; Evolution; Allometry; Mammal

Introduction

"Scaling is interesting because, aside from natural selection, it is one of the few laws we really have in biology," said John Gittleman in the New York Times on January 12, 1999. The appeal of using scaling relationships in understanding, modelling and predicting biological phenomena is documented in various books dedicated to this approach (Peters 1983; Calder 1984; Schmidt-Nielsen 1984; Bonner 2006; Sibly et al. 2012). In most simple terms, the scaling approach relies on statistical relationships between body mass (M) and other measurements that usually follows the relationship $y = a\ M^b$. Various of such relationships have been linked with each other, in particular with respect to the magnitude of the exponent. In general terms, it is often assumed that the exponent is a multiple of ¼, such as ¾ for metabolic rates (or energy requirements or food intake), -¼ for physiological and ecological rates (such as heart rate, population growth rate), or ¼ for biological times (such as digesta retention time, gestation period, or longevity) (Brown et al. 2012). A causal link between these relationships is a fundamental assumption of the Metabolic Theory of Ecology (MTE; Brown et al. 2004).

Testing whether certain physiological or ecological measures actually follow the predicted scaling relationships is therefore considered one way to test the MTE, and a large number of such tests have been published (reviewed in Sibly et al. 2012). These tests are usually performed on datasets that comprise large numbers of species. Because these data are not independent, but may be influenced by – amongst several factors – phylogenetic relationships between the species, the use of statistical methods that also account for the phylogenetic structure of the dataset has been recommended for a long time (Harvey and Pagel 1991; White et al. 2012). Nevertheless, it is sometimes assumed that if datasets are large enough, results from statistics without accounting for phylogeny, such as in Ordinary Least Squares (OLS), are similar to those from statistics that do account for the phylogenetic structure, such

as Phylogenetic Generalized Least Squares (PGLS) (White et al. 2012). For example, the allometric exponent for the scaling of basal metabolic rate with body mass is similar in a large dataset of mammals, at 0.72 without and 0.73 with accounting for phylogeny (Müller et al. 2012). Actually, a significant result in conventional statistics is often considered a prerogative for further considerations, because without such significance, the universality of the relationship can usually not be postulated in the first place (but see e.g. Zerbe et al. 2012 for an exception).

Of course, uniformity between OLS and PGLS needs not always be the case. Depending on the structure of the data, several different combinations of results are possible (Fig. 1). If there is a similar significant relationship in PGLS as in OLS, the described pattern appears to be universal (Fig. 1a). If both OLS and PGLS are significant and yield the same allometric relationship, this can also occur because of (apparently random) variation in this relationship at lower taxonomic levels (Fig. 1b). Because the scaling of basal metabolic rate differs between individual mammal taxa (Sieg et al. 2009; White et al. 2009; Capellini et al. 2010) yet the scaling is similar with and without accounting for phylogeny in the entire mammal dataset (Müller et al. 2012), the example in Fig. 1b likely reflects, qualitatively, the situation for the scaling of basal metabolism. In this case, one might argue that the basic pattern is universal, but may be modified to certain degrees by characteristics of certain taxonomic groups. The magnitude of the individual deviations will then determine whether universality can still be accepted for the pattern. The situation changes, however, if a Type I error occurs, which can come in two different versions: On the one hand, the underlying pattern might just be evident because of the array of taxa, with no relationship among the more closely related species – a case with a significant result in OLS but no significance in PGLS (Fig. 1c). Such a result may indicate that the assumed pattern does exist at the level of the basal node (and therefore could only be called a ‘pattern’ if it was found at this level if additional clades at the

basal node were included in the analysis), but not among more closely related extant species. On the other hand, in a similar scenario, both OLS and PGLS may yield significant results, but with different estimates for the exponent (Fig. 1d). Such a result again indicates that the assumed relationship systematically changes its pattern, depending on the taxonomic level of analysis. Both such results (Fig. 1cd) cannot be reconciled with the assumption of a universal physiological pattern, but must lead to hypotheses that link the occurrence or shape of the pattern to taxonomic levels of organisation.

When simply analysing for the existence of a relationship between two variables – as often done in studies that test for convergence, for example by comparing the magnitude of a morphological or physiological adaptation to an environmental proxy -, a situation as indicated in Fig. 1d is of little relevance. This is because the exact magnitude of the convergent adaptation is not in question, just its existence. Only a situation as indicated in Fig. 1c will refute the hypothesis of convergence. In such situations – when testing 'only' for convergence -, results from phylogenetically informed statistics mostly confirm those of conventional approaches (e.g. Ricklefs and Starck 1996; Carvalho et al. 2006). If, however, scaling laws are investigated, it is not only the existence of a significant relationship between two parameters *per se*, but the exact pattern of that relationship that is of interest.

Here, we investigate the scaling pattern of gestation period with body mass in eutherian mammals, providing an example of how analyses that account for the phylogenetic structure of datasets can lead to alternative interpretations of evolutionary trends. Eutherian gestation periods appear as a particularly suited test case for scaling analysed in a phylogenetic context. On the one hand, they are hypothesized to scale to $M^{0.25}$ (Western 1979; Lindstedt and Calder 1981; Brown et al. 2012), a pattern more or less confirmed in analyses that do not account for phylogeny, where the scaling pattern varies between $M^{0.17}$ to $M^{0.24}$, depending on the species

composition of the dataset such as the inclusion or exclusion of bats (Kihlström 1972; Hamilton et al. 2011; Sibly 2012). On the other hand, there is evidence that closely related organisms do not follow such a scaling pattern. In domestic animals, it is well-known that the gestation period does not scale with *M* intraspecifically among different-sized breeds: a 5 kg dog breed has the same gestation period as a 50 kg dog breed (Kirkwood 1985), gestation period is nearly constant across rabbit breeds ranging from 2 to 6 kg (King Wilson and Dudley 1952), a Shetland pony has the same gestation period as a Friesian horse (Bos and Van der Mey 1980), and dwarf breeds have similar gestation periods as normal-sized breeds in goats (Amoah et al. 1996), sheep (Jollans 1960; Bradford et al. 1972) and cattle (Crew 1923; Andersen and Plum 1965). Additionally, gestation period showed little interspecific scaling within the genera *Felis*, *Panthera*, *Peromyscus*, *Cercopithecus* and *Macaca*, respectively (Martin and MacLarnon 1985) or within bats (Jones and Purvis 1997), and is generally lower within altricial and precocial species if these are analysed separately, and in many mammalian clades than in datasets covering multiple clades (Martin et al. 2005), for example in primates (Dubman et al. 2012). Moreover, there seems to be some indication that in ruminants, secondarily dwarfed species such as the anoa (*Bubalus depressicornis*) retain the (relatively long) gestation periods of their larger relatives (Müller et al. 2011). Actually, the comparatively long gestation periods of some dwarf antelopes have been used to argue that these species represent cases of secondary dwarfism (Geist 1966). Therefore, we predict that gestation period shows a difference in the interspecific scaling pattern as analysed by OLS and PGLS.

Methods

Data on extant eutherian body mass and length of the gestation period were taken from the PanTheria database (Jones et al. 2009). Because of the well-known difference in gestation period length between altricial and precocial species (Martin and MacLarnon 1985; Martin et

al. 2005) but absence of this information for many species in the dataset, we used the number of offspring per litter as a surrogate proxy (Müller et al. 2012; Müller et al. 2013), classifying species as those with one, and those with more than one offspring per litter. The data were linked to a supertree of extant mammals (Bininda-Emonds et al. 2007; 2008), resulting in a final dataset of 1214 species from 20 orders, with an evident difference in the scaling between eutherians with one and more than one offspring per litter (Fig. 2).

We tested the relationship between log-transformed *M* as the independent, and log-transformed gestation period length as the dependent variable, using Ordinary Least Squares (OLS) and Phylogenetic Generalized Least Squares (PGLS) in the whole dataset and in taxonomic subsets. Littersize was included as a two-level (litter size = 1 and litter size > 1) cofactor in several analyses.

While PGLS was used with Pagel's λ, the character describing the phylogenetic signal (Pagel 1999; Revell 2010), fixed at λ=1 for the series of analyses in various randomised data subsets (assuming Brownian motion), we additionally estimated λ using maximum likelihood for the whole dataset and several data subsets (Revell 2010). λ can vary between 0 (no phylogenetic signal) and 1 (the observed pattern is predicted by the phylogeny; similarity among species scales in proportion to their shared evolutionary time) (Pagel 1999; Freckleton et al. 2002).

To test whether the use of PGLS was adequate, we compared the fit of OLS and PGLS approaches for the complete dataset using their respective model log-likelihoods, as compared by the likelihood ratio test using a chi-square distribution (Pagel 1999; Freckleton et al. 2002). Statistical tests were performed in *R* 2.15.0 (Team 2011) using the packages ape (Paradis et al. 2004), caper (Orme et al. 2010), nlme (Pinheiro et al. 2011) and phytools (Revell 2012).

Results

174 The scaling in the overall dataset differed between OLS and PGLS (Table 1); the scaling
175 exponent was distinctly higher in OLS (0.18-0.20) than in PGLS (0.07-0.10). Notably, even in
176 OLS the confidence interval of the scaling exponent did not include 0.25. For the overall
177 dataset, λ was estimated at 0.992 (95% confidence interval 0.982-0.998), indicating that more
178 closely related species showed similar trait values, and that the OLS model therefore is
179 inadequate for this dataset. The log-likelihoods in the default PGLS model where λ=1
180 (1208.426) and with actually estimated λ of 0.99 (1214.058) were both distinctively higher
181 than the one for the OLS model (-159.438), and a chi-square comparison between the log-
182 likelihoods of OLS and PGLS correspondingly indicated a highly significant difference
183 ($p<0.001$), indicating that the PGLS models provided a better data fit.
184 Because both litter size and the interaction of litter size with *M* were, as expected from the
185 data plot (Fig. 2), significant in OLS (t=-25.39, $p<0.001$ and t=8.17, $p<0.001$, respectively),
186 the two groups of species were additionally analysed separately. For species with one
187 offspring per litter, scaling exponents were numerically similar in OLS and in PGLS,
188 although they did not quite overlap at 95% confidence (0.097-0.110, and 0.061-0.095,
189 respectively); for species with more than one offspring per litter, the situation was similar as
190 for the whole dataset, with an exponent distinctly higher in OLS (0.15-0.18) than in PGLS
191 (0.06-0.10). Interestingly, the scaling exponent in PGLS was similar in the whole dataset and
192 in the subsets by littersize category, indicating that littersize categories were reflected in the
193 phylogenetic structure of the dataset (Table 1). In all these cases, the 95%CI for the intercept
194 of the allometric regression equation was smaller in OLS than in PGLS (Table 1).
195 In order to demonstrate how results in OLS and PGLS differed, we analysed random subsets
196 that allowed only one species per order (100 subsets with 20 species each) and per family
197 (100 subsets with 105 species each), and one dataset each with an increasing number of
198 species allowed per genus (from a maximum of 1 up to a maximum of 15 species per genus).
199 While we are aware that the distinction of these different taxonomic levels implies a

200 (historical) arbitrariness, so that e.g. a ‘family’ in one order might rather correspond to a
201 ‘genus’ in another order, this approach nevertheless guarantees a successive increase in the
202 number of more closely related species. The results indicate that both the intercept and the
203 scaling exponent differ between OLS and PGLS at the genus level (Fig. 3); at this level, the
204 intercept is higher (and has a larger 95% confidence interval) in PGLS, and the scaling
205 exponent in PGLS was about half the magnitude as that in OLS. An exponent of 0.25 was
206 only included at the order level in both OLS and PGLS. In PGLS, allowing a larger number of
207 more closely related species led to an increase of the intercept and a decline of the scaling
208 exponent.
209 In orders with more than 10 species, a similar pattern as in the overall dataset was evident
210 (Table 1). Either scaling exponents were lower in PGLS than in OLS, or they were similar,
211 and low, at around 0.10. In the three families that had more than 50 species, the Bovidae,
212 Muridae and Sciuridae, the results followed a similar pattern (Table 1). Generally, results
213 were similar whether λ was fixed or estimated in the PGLS procedure. The only two orders in
214 which λ was not significantly different from zero were the lagomorphs and the perissodactyls,
215 the two taxonomic subsets with the lowest sample sizes (Table 1).
216
217 **Discussion**
218 Eutherian gestation periods are a prime example where analyses accounting for the
219 phylogenetic structure of the dataset yield relevantly different results compared to
220 conventional analyses, a fact that in itself cannot be reconciled with a universal scaling
221 pattern. In particular, the results demonstrate that even though a quarter-power scaling of
222 gestation period might be reached in eutherians at the order level, such a scaling cannot be
223 supported at lower taxonomic levels, where gestation period scales at about $M^{0.1}$ or lower
224 among closely related extant species. Within closely related taxa, shifts in body mass are thus

not associated with the large shifts in gestation period length that one would expect if quarter-power scaling was operating, or that one can observe between not so closely related taxa.

Rather than representing a universal scaling characteristic, the high scaling at high taxonomic levels in OLS appears to be particularly linked to altriciality (or the production of several offspring per litter). Among the precocial species alone, differences in the scaling exponent in OLS and PGLS were slight (Table 1). As noted previously, altriciality is associated with shorter gestation periods and other life history and physiologic characteristics, such as higher rates of population increase (Hennemann 1984; Müller et al. 2012). Most importantly for allometric analyses, altriciality is mostly constrained to the lower body mass range of mammals, possibly due to reasons unrelated to metabolism (Müller et al. 2012), whereas precociality or the production of a single offspring per litter occurs across the whole body mass spectrum. In large comparative datasets, this systematic bias can generate different scaling effects that should be controlled for. Interestingly, such a control can be performed not only by testing for scaling within different reproductive groups or groups that differ in other fundamental biological traits, but in the case of gestation period also by using PGLS in a large dataset, indicating that the distribution of these traits is reflected in mammalian phylogeny and that the scaling pattern is not fundamentally different between the two reproductive groups.

The results suggest that evolutionary shifts in body mass are related to different modifications in organismal development at different taxonomic levels. At lower taxonomic levels, interspecific changes in body mass are not primarily achieved by changes in development time, which shows only a low scaling exponent, but must therefore be realized mainly by altering development speed. Unravelling the physiological mechanisms behind such changes in development speed appears a promising area of research. While such changes are most

likely linked to differences in growth hormone regulation (Bernstein 2010), it is unlikely that they have a direct link to metabolism. Based on the findings in different domestic dogs breeds mentioned in the Introduction, Kirkwood (1985) suggested that gestation is controlled by a 'clock' that is independent of body size and thus metabolic rate; our results suggest that this holds true not only during shorter evolutionary time scales like artificial selection process of domestication, but also during more recent radiation processes within a linage, i.e. at the genus level. Other arguments also suggest that gestation period is to a large degree independent from metabolism. Whereas metabolic scaling is similar in a large eutherian dataset whether analysed without or with accounting for phylogeny (Müller et al. 2012), this does not apply to gestation period (Fig. 3). Whereas metabolic scaling differs between eutherian littersize categories even when controlling for phylogeny (Müller et al. 2012), no difference in the scaling of gestation period was evident between the two groups in this study (Table 1). Additionally, the scaling exponent of gestation period in the overall dataset or the littersize category subsets – 0.06-0.10 – cannot be directly linked to the exponents for metabolic scaling (about 0.67-0.75) in the form of reciprocal quarter-power pattern as proposed in the Metabolic Theory of Ecology. Differences in intrauterine developmental speed between closely related species of different sizes, therefore, are unlikely directly linked to differences in metabolism one would derive from these differences in body size.

These findings are particularly relevant for case studies in dwarfism and gigantism. As already mentioned, the anoas – dwarf buffaloes of the Sulawesi island – have longer gestation periods, and higher maximum longevities, than expected from their body size (Müller et al. 2011). This suggests that using gestation period length to detect cases of secondary dwarfism, as originally suggested by Geist (1966), is a reasonable approach. It also suggests that dwarfism is achieved by a reduction in development speed, i.e. reduced growth rates – a hypothesis supported by analyses of bones and teeth in a fossil dwarf bovid from Majorca

(Köhler and Moyà-Solà 2009; Jordana and Köhler 2011). In the debate on the life history consequences of insular dwarfism, however, the opposite – a faster life history –, has repeatedly been suggested (Raia et al. 2003; Palombo 2007; Meiri and Raia 2010). Studies that link insular dwarfism to a shift towards a faster life history often rely on the reconstruction of life history traits using allometry. For example, Roth (1990; 1992), Raia et al. (2003) or Palombo (2007) explicitly use quarter-power scaling to reconstruct the life histories of insular dwarf elephants – an approach not supported by the results of our study. In contrast, findings on comparatively fast tooth growth in large, and slow tooth growth in small fossil elephant species (Dirks et al. 2012) might support the hypothesis on variation in growth rate resulting from our study. Broadening the animal spectrum beyond mammals, an increase in growth rate (rather than an increase in development time) in gigantism and a decrease in growth rate (rather than a decrease in development time) in insular dwarfism have also been suggested, based on histological investigations of bones, in various dinosaurs (Sander et al. 2004; Sander et al. 2006; Benton et al. 2010; Stein et al. 2010). The pattern that development time is rather constant but development speed varies between closely related species of different body size may thus emerge as a more general rule that encompasses various animal clades in future studies. Importantly, this approach focuses on a universal physiological mechanism of body size variation in closely related animals, rather than a variation in extrinsic factors such as mortality and resource availability (e.g. Jordana et al. 2012).

Apart from the low scaling of gestation period among closely related species, the general variation at higher taxonomic levels needs to be accounted for; actually, this is reflected in the higher variation of the intercept of the allometric equation in PGLS. Note that the 95% confidence interval for the intercept in the overall dataset is much larger in PGLS than in OLS (Table 1). Differentiation of taxonomic lineages based on differences in the intercept correspond to the finding of Egset et al. (2012), who showed in a selection experiment that it

303 was easier to change the intercept than the slope of an allometric relationship for a
304 morphological trait. If taxonomic lineage differentiatiation is accompanied by changes in the
305 intercept, this could induce a different conceptual approach to the interpretation of
306 comparative datasets. A conceptual decision in dealing with allometries in comparative
307 datasets is whether one assumes the resulting relationship to represent some fundamental rule
308 or natural law that indicates a physiological constraint. The scaling of basal metabolic rate
309 with body mass has reached such a status (whether this is correct or not), and even if the exact
310 nature of the exponent or the groups of animals to which one common exponent applies may
311 still be under debate, the existence of scaling rules for metabolism are usually not disputed
312 (Müller et al. 2012) (but see e.g. Kozlowski and Weiner 1997). By contrast, allometric
313 relationships originating from comparative data collections can also be considered as
314 snapshots in evolutionary time, which will vary depending on the species included, and thus
315 depend on the sequence of species radiations and extinctions that occurred before the picture
316 was taken. A physiological innovation may allow a taxonomic lineage to change the intercept
317 of the allometric relationship; once this lineage has evolved or diversified, its contribution
318 will shift the allometric relationship of the overall dataset in the direction of this new
319 intercept. For example, Fritz et al. (2009) investigated the allometric relationship between
320 body size and faecal particle size in herbivores, and observed that taxonomic groups with
321 more recent radiations (e.g. equids as compared to rhinoceroses, or ruminants as compared to
322 equids) achieved finer faecal particles. Prior to the diversification of ruminants, such an
323 allometry would therefore have shown a different picture. The allometric relationship between
324 body mass and faecal particle size therefore does not represent a universal law, but is subject
325 to modifications by morphophysiological innovations.
326
327 This concept can also be applied to the present study. One evident difference between equids
328 and ruminants are the shorter gestation periods achieved by ruminants at similar sizes (Zerbe

et al. 2012). Within the artiodactyla, a similar difference in the gestation period length between camelids and giraffids on the one hand – both groups with a low diversity of extant species – and the bovids on the other hand, with their high species diversity and shorter gestation periods, is also striking (Müller et al. 2011). Yet why a calf (cattle) attains an apparently similar level of precociality after app. 280 days of gestation as a foal (horse) after app. 340 days, a dromedary (*Camelus dromedarius*) newborn after app. 390 days, or an okapi (*Okapia johnstoni*) newborn after app. 440 days remains, to our knowledge, unexplored so far. We only know that these differences cannot be explained by differences in body size. Thus, rather than considering life history parameters such as gestation period as characters subject to static scaling relationships mainly determined by body size, they can also be understood as variables by which different taxonomic lineages may compete for niche space, for example by reducing gestation periods in order to optimally use seasonal habitats (Zerbe et al. 2012) or to reduce generation times (and thereby increase fecundity). The differences in the intercept *a* in Table 1 between taxonomic groups, and the difference in the scaling between OLS and PGLS suggests that by differing in organismal design (and hence phylogenetic affiliation), animals can achieve variation in the characteristic under investigation (here: gestation period) that they could not achieve by simply varying body size. To investigate the mechanism behind such fundamental differences in organismal design appears a promising area of research that might expand our understanding for the historical succession of radiations of different taxonomic groups competing for shared resources. For example, we might predict that in general, the taxa with the potentially more advanced characteristic – here, the shorter gestation period (at a similar level of development of the offspring) - should be the more diverse, as when comparing bovids against giraffids, camelids, or perissodactyls (note that the latter three groups all once comprised a higher diversity and covered broader body size ranges in the fossil record). For zebras, for example, the comparatively longer generation times than sympatric ruminants have been implied as a

355 factor limiting their abundance (Grange et al. 2004). Because equid gestation periods exceed
356 a year, it is difficult for equids to synchronize births with a seasonal pattern without losing
357 reproductive opportunity, a factor that may make them both more susceptible to predation
358 (Grange and Duncan 2006) and less adapted to highly seasonal environments (Zerbe et al.
359 2012). How the surviving species of those taxa with the lesser advantageous characteristic can
360 defend their extant ecological niches must then depend on other factors.
361
362 We conclude that gestation period, analysed by PGLS, shows a relevantly lower allometric
363 scaling with body mass than often communicated. This could suggest that interspecific
364 changes in body mass are not predominantly effected by changes in development time, but
365 (also) by changes in intrauterine development speed during evolution. A direct link between
366 the scaling of gestation period and that of metabolism is not evident. The uneven distribution
367 of species across the body mass spectrum when classified by some proxy of littersize or
368 developmental mode (altricial/precocial) will lead to artefacts in scaling relationships that
369 need to be accounted for. Because the dichotomy in developmental mode is partially reflected
370 in the phylogeny of mammals, PGLS-based analyses will often correct for this effect, unless
371 the scaling itself varies between altricial and precocial species. Larger datasets, with larger
372 numbers of closely related species, will make potential differences between OLS and PGLS
373 analyses more evident. If other life history variables, such as longevity, time to weaning, time
374 to first reproduction etc. would show similar deviations from traditionally assumed scaling
375 patterns when analysed by PGLS, a re-assessment of theories that rely on the universality of
376 quarter-power scaling, and that link life history directly to metabolism, would be warranted.
377
378 **Acknowledgements**
379 MC and DC were supported within the DFG Research Unit 533 “Biology of the Sauropod
380 Dinosaurs: The Evolution of Gigantism”, of which this is publication no. xx.

References

Amoah EA, Gelaye S, Guthrie P, Rexroad CE (1996) Breeding season and aspects of reproduction of female goats. Journal of Animal Science 74:723-728

Andersen H, Plum M (1965) Gestation length and birth weight in cattle and buffaloes: a review. Journal of Dairy Science 48:1224-1235

Baker RH (2002) Comparative methods. In: DeSalle R, Giribet G, Wheeler W (eds) Techniques in molecular systematics and evolution. Birkhäuser Verlag, Basel, pp 146-161

Benton MJ, Csiki Z, Grigorescu D, Redelstorff R, Sander PM, Stein K, Weishampel DB (2010) Dinosaurs and the island rule: the dwarfed dinosaurs from Hateg Island. Palaeogeography, Palaeoclimatology, Palaeoecology 293:438–454

Bernstein RM (2010) The big and the small of it: how body size evolves. Yearbook of Physical Anthropology 53:46-62

Bininda-Emonds ORP, Cardillo M, Jones KE, MacPhee RDE, Beck RMD, Grenyer R, Price SA, Vos RA, Gittleman JL, Purvis A (2007) The delayed rise of present-day mammals. Nature 446:507-512

Bininda-Emonds ORP, Cardillo M, Jones KE, MacPhee RDE, Beck RMD, Grenyer R, Price SA, Vos RA, Gittleman JL, Purvis A (2008) Corrigendum: The delayed rise of present-day mammals. Nature 456:274

Bonner JT (2006) Why size matters. Princeton University Press, Princeton

Bos H, Van der Mey GJW (1980) Length of gestation periods of horses and ponies belonging to different breeds. Livestock Production Science 7:181-187

Bradford GE, Hart R, Quirke JF, Land RB (1972) Genetic control of the duration of gestation in sheep. Journal of Reproduction and Fertility 30:459-463

Brown JH, Gillooly JF, Allen AP, Savage VM, West GB (2004) Towards a metabolic theory of ecology. Ecology 85:1771–1789

Brown JH, Sibly RM, Kodric-Brown A (2012) Introduction: metabolism as the basis for a theoretical unification of ecology. In: Sibly RM, Brown JH, Kodric-Brown A (eds) Metabolic ecology A scaling approach. Wiley-Blackwell, Chichester, UK, pp 1-6

Calder WA (1984) Size, function, and life history. Harvard University Press, Cambridge, MA

Capellini I, Venditi C, Barton RA (2010) Phylogeny and metabolic scaling in mammals. Ecology 91:2783–2793

Carvalho P, Diniz-Filho J, Bini L (2006) Factors influencing changes in trait correlations across species after using phylogenetic independent contrasts. Evolutionary Ecology 20:591-602

Crew FAE (1923) The significance of an achondroplasia-like condition met with in cattle. Proceedings of the Royal Society B 95:228-255

Dirks W, Bromage TG, Agenbroad LD (2012) The duration and rate of molar plate formation in *Palaeoloxodon cypriotes* and *Mammuthus columbi* from dental histology. Quaternary International 255:79-85

Dubman E, Collard M, Mooers AØ (2012) Evidence that gestation duration and lactation duration are coupled traits in primates. Biology Letters 8:998-1001

Egset CK, ansen TF, Le Rouzic A, Bolstad GH, Rosenqvist G, Pélabon C (2012) Artificial selection on allometry: change in elevation but not slope. Journal of Evolutionary Biology 25:938–948

Freckleton RP, Harvey PH, Pagel M (2002) Phylogenetic analysis and comparative data: a test and review of evidence. American Naturalist 160:712–726

430 Fritz J, Hummel J, Kienzle E, Arnold C, Nunn C, Clauss M (2009) Comparative chewing
431 efficiency in mammalian herbivores. Oikos 118:1623-1632
432 Geist V (1966) The evolution of horn-like organs. Behaviour 27:175-214
433 Grange S, Duncan P, Gaillard J-M, Sinclair ARE, Gogan PJP, Packer C, Heribert H, East M
434 (2004) What limits the Serengeti zebra population? Oecologia 140:523-532
435 Grange S, Duncan P (2006) Bottom-up and top-down processes in African ungulate
436 communities: resources and predation acting on the relative abundance of zebra and
437 grazing bovids. Ecography 29:899–907
438 Hamilton MJ, Davidson AD, Sibly RM, Brown JH (2011) Universal scaling of production
439 rates across mammalian lineages. Proceedings of the Royal Society B 278:560-566
440 Harvey PH, Pagel MD (1991) The comparative method in evolutionary biology. Oxford
441 University Press, Oxford
442 Hennemann WW (1984) Intrinsic rates of natural increase of altricial and precocial eutherian
443 mammals: the potential price of precociality. Oikos 43:363-368
444 Jollans JL (1960) A study of the West African dwarf sheep in the closed forest zone of
445 Ashanti. West African Journal of Biological Chemistry 3:74-80
446 Jones KE, Purvis A (1997) An optimum body size for mammals? Comparative evidence from
447 bats. Functional Ecology 11:751–756
448 Jones KE, Bielby J, Cardillo M, Fritz SA, O'Dell J, Orme CD, Safi K, Sechrest W, Boakes
449 EH, Carbone C, Connolly C, Cutts MJ, Foster JK, Grenyer R, Habib M, Plaster CA,
450 Price SA, Rigby EA, Rist J, Teacher A, Bininda-Emonds ORP, Gittleman JL, Mace
451 GM, Purvis A (2009) PanTHERIA: a species-level database of life history, ecology,
452 and geography of extant and recently extinct mammals. Ecology 90:2648 (Ecological
453 Archives E2090-2184)
454 Jordana X, Köhler M (2011) Enamel microstructure in the fossil bovid *Myotragus balearicus*
455 (Majorca, Spain): implications for life-history evolution of dwarf mammals in insular
456 ecosystems. Palaeogeography, Palaeoclimatology, Palaeoecology 300:59-66
457 Jordana X, Marín-Moratalla N, DeMiguel D, Kaiser TM, Köhler M (2012) Evidence of
458 correlated evolution of hypsodonty and exceptional longevity in endemic insular
459 mammals. Proceedings of the Royal Society B online
460 Kihlström JE (1972) Period of gestation and body weight in some placental mammals.
461 Comparative Biochemistry and Physiology A 43:673-679
462 King Wilson W, Dudley FJ (1952) The duration of gestation in rabbit breeds and crosses.
463 Journal of Genetics 50:384-391
464 Kirkwood JK (1985) The influence of size on the biology of the dog. Journal of Small Animal
465 Practice 26:97-110
466 Köhler M, Moyà-Solà S (2009) Physiological and life history strategies of a fossil large
467 mammal in a resource-limited environment. Proceedings of the National Academy of
468 Science 106:20354-20358
469 Kozlowski J, Weiner J (1997) Interspecific allometries are by-products of body size
470 optimization. American Naturalist 149:352–380
471 Lindstedt SL, Calder WAI (1981) Body size, physiological time and longevity of
472 homeothermic animals. Quarterly Review of Biology 56:1-16
473 Martin RD, MacLarnon AM (1985) Gestation period, neonatal size and maternal investment
474 in placental mammals. Nature 313:220-223
475 Martin RD, Genoud M, Hemelrijk CK (2005) Problems of allometric scaling analysis:
476 examples from mammalian reproductive biology. Journal of Experimental Biology
477 208:1731-1747
478 Meiri S, Raia P (2010) Reptilian all the way? Proceedings of the National Academy of
479 Science 107:E27

480 Müller DWH, Zerbe P, Codron D, Clauss M, Hatt J-M (2011) A long life among ruminants:
481 giraffids and other special cases. Schweizer Archiv für Tierheilkunde 153:515-519
482 Müller DWH, Codron D, Werner J, Fritz J, Hummel J, Griebeler EM, Clauss M (2012)
483 Dichotomy of eutherian reproduction and metabolism. Oikos 121:102-115
484 Müller DWH, Codron D, Meloro C, Munn A, Schwarm A, Hummel J, Clauss M (2013)
485 Assessing the Jarman-Bell Principle: scaling of intake, digestibility, retention time and
486 gut fill with body mass in mammalian herbivores. Comparative Biochemistry and
487 Physiology A 164:129–140
488 Nunn CL (2011) The comparative approach in evolutionary anthropology and biology.
489 University of Chicago Press, Chicago
490 Orme D, Freckleton R, Thomas G, Petzoldt T, Fritz S, Isaac N (2010) Caper: comparative
491 analyses of phylogenetics and evolution in R. R package version 04/r71 See
492 http://wwwR-ForgeR-projectorg/projects/caper/
493 Pagel M (1999) Inferring the historical patterns of biological evolution. Nature 401:877-884
494 Palombo MR (2007) How can endemic proboscideans help us understand the "island rule"? A
495 case study of Mediterranean islands. Quaternary International 169/170:105-124
496 Paradis E, Claude J, Strimmer K (2004) APE: analyses of phylogenetics and evolution in R
497 language. Bioinformatics 20:289–290
498 Peters RH (1983) The ecological implications of body size. Cambridge University Press,
499 Cambridge
500 Pinheiro J, Bates D, DebRoy S, Sarkar D, R Development Core Team (2011) Nlme: linear
501 and nonlinear mixed effects models. R package version 3 1–102 Available at
502 http://cranrprojectorg/web/packages/nlme/citationhtml
503 Raia P, Barbera C, Conte M (2003) The fast life of a dwarfed giant. Evolutionary Ecology
504 15:293-312
505 Revell LJ (2010) Phylogenetic signal and linear regression on species data. Methods in
506 Ecology and Evolution 1:319-329
507 Revell LJ (2012) Phytools: an R package for phylogenetic comparative biology (and other
508 things). Methods in Ecology and Evolution 3:217–223
509 Ricklefs RE, Starck JM (1996) Applications of phylogenetically independent contrasts: a
510 mixed progress report. Oikos 77:167-172
511 Roth VL (1990) Island dwarf elephants: a case study in body mass estimatio and ecological
512 inference. In: Damuth J, MacFadden JB (eds) Body size in mammalian paleobiology.
513 Cambridge University Press, New York, pp 51-179
514 Roth VL (1992) Inferences from allometry and fossils: dwarfing elephant on islands. In:
515 Futuyma D, Antonovics J (eds) Oxford surveys in evolutionary biology, vol 8. Oxford
516 University Press, New York, pp 259-288
517 Sander PM, Klein N, Buffetaut E, Cuny G, Suteethorn V, Le Loeuff J (2004) Adaptive
518 radiation in sauropod dinosaurs: bone histology indicates rapid evolution of giant body
519 size through acceleration. Organisms, Diversity & Evolution 4:165-173
520 Sander PM, Mateus O, Laven T, Knötschke N (2006) Bone histology indicates insular
521 dwarfism in a new Late Jurassic sauropod dinosaur. Nature 441:739-741
522 Schmidt-Nielsen K (1984) Scaling: Why is animal size so important? Cambridge University
523 Press, Cambridge
524 Sibly RM (2012) Life history. In: Sibly RM, Brown JH, Kodric-Brown A (eds) Metabolic
525 ecology A scaling approach. Wiley-Blackwell, Chichester, UK, pp 57-66
526 Sibly RM, Brown JH, Kodric-Brown A (eds) (2012) Metabolic ecology. A scaling approach.
527 Wiley-Blackwell, Chichester, UK
528 Sieg AE, O'Connor MP, McNair JN, Grant BW, Agosta SJ, Dunham AE (2009) Mammalian
529 metabolic allometry: do intraspecific variation, phylogeny, and regression models
530 matter? American Naturalist 174:720-733

Stein K, Csiki Z, Curry Rogers K, Weishampel DB, Redelstorff R, Carballido JL, Sander PM (2010) Small body size and extreme cortical bone remodeling indicate phyletic dwarfism in *Magyarosaurus dacus* (Sauropoda: Titanosauria). Proceedings of the National Academy of Science 107:9258-9263

Team RDC (2011) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria ISBN 3-900051-07-0, URL http://wwwR-projectorg/

Western D (1979) Size, life history and ecology in mammals. African Journal of Ecology 17:185-204

White CR, Blackburn TM, Seymour RS (2009) Phylogenetically informed analysis of the allometry of mammalian basal metabolic rate supports neither geometric nor quarter-power scaling. Evolution 63:2658-2667

White EP, Xiao X, Isaac NJB, Sibly RM (2012) Methodological tools. In: Sibly RM, Brown JH, Kodric-Brown A (eds) Metabolic ecology A scaling approach. Wiley-Blackwell, Chichester, UK, pp 9-20

Zerbe P, Clauss M, Codron D, Bingaman Lackey L, Rensch E, Streich WJ, Hatt J-M, Müller DHW (2012) Reproductive seasonality in captive wild ruminants: implications for biogeographical adaptation, photoperiodic control, and life history. Biological Reviews 87:965-990

Table 1. Parameter estimates for the scaling of eutherian gestation period (G, in days) with body mass (M, in grams) according to $G = aM^b$ as analysed in various taxonomic and functional subsets of the dataset, analysed by Ordinary Least Squares (OLS, i.e. without accounting for the phylogenetic structure of the data) and by Phylogenetic Generalized Least Squares (PGLS, i.e. with accounting for the phylogenetic structure of the data, with λ either fixed to 1 or estimated by maximum likelihood). Differences between OLS and PGLS marked by grey shading.

Dataset	*n*	Statistic	λ	*a*	95%CI	*t*	*p*	*b*	95%CI	*t*	*p*
All species	1214	OLS	(0)	21.5	19.9-23.3	75.11	<0.001	0.19	0.18-0.20	36.27	<0.001
		PGLS	1	53.8	42.1-68.9	31.65	<0.001	0.08	0.07-0.10	12.72	<0.001
			0.99**	52.4	41.3-66.3	32.83	<0.001	0.09	0.08-0.10	13.34	<0.001
1offspring	541	OLS	(0)	73.5	69.2-78.2	136.78	<0.001	0.10	0.10-0.11	30.99	<0.001
		PGLS	1	77.2	59.2-100.7	32.13	<0.001	0.08	0.06-0.10	8.99	<0.001
			0.98**	74.4	58.3-95.0	34.58	<0.001	0.08	0.07-0.10	9.81	<0.001
>1offspring	673	OLS	(0)	16.7	15.3-18.3	61.42	<0.001	0.17	0.15-0.18	22.57	<0.001
		PGLS	1	41.7	31.9-54.6	27.19	<0.001	0.08	0.06-0.10	8.78	<0.001
			1.00*	41.7	31.9-54.6	27.19	<0.001	0.08	0.06-0.10	8.78	<0.001
Cetartiodactyla	218	OLS	(0)	68.1	55.5-83.6	40.29	<0.001	0.11	0.09-0.12	11.82	<0.001
		PGLS	1	101.3	72.3-142.0	26.80	<0.001	0.07	0.06-0.09	7.59	<0.001
			0.97**	96.6	71.4-130.6	29.67	<0.001	0.08	0.06-0.10	8.41	<0.001
Artiodactyla	175	OLS	(0)	66.3	49.0-89.7	27.26	<0.001	0.10	0.08-0.13	7.44	<0.001
		PGLS	1	93.0	62.7-137.8	22.57	<0.001	0.07	0.05-0.10	6.06	<0.001
			0.98**	88.3	61.6-126.7	24.36	<0.001	0.08	0.06-0.10	6.64	<0.001
Cetacea	43	OLS	(0)	282.3	225.5-353.4	49.26	<0.001	0.02	0.00-0.03	2.07	0.045
		PGLS	1	136.4	96.9-192.1	28.16	<0.001	0.07	0.04-0.09	5.61	<0.001
			1.00*	136.4	96.9-192.1	28.16	<0.001	0.07	0.04-0.09	5.61	<0.001
Carnivora	173	OLS	(0)	7.1	5.5-9.2	15.17	<0.001	0.28	0.25-0.30	19.94	<0.001
		PGLS	1	35.5	19.9-63.4	12.07	<0.001	0.11	0.07-0.15	5.78	<0.001
			1.00*	35.5	19.9-63.4	12.07	<0.001	0.11	0.07-0.15	5.78	<0.001
Chiroptera	151	OLS	(0)	70.6	60.4-82.7	53.02	<0.001	0.13	0.08-0.18	5.36	<0.001
		PGLS	1	117.8	78.2-177.5	22.79	<0.001	0.00	-0.06-0.07	0.09	0.930
			0.95**	116.2	80.8-167.1	25.66	<0.001	0.01	-0.06-0.07	0.20	0.842
Insectivora	47	OLS	(0)	18.7	16.2-21.6	39.55	<0.001	0.13	0.10-0.17	7.23	<0.001
		PGLS	1	26.0	17.8-37.9	16.95	<0.001	0.07	0.02-0.13	2.47	0.018
			0.58**	23.4	18.3-29.9	25.29	<0.001	0.09	0.04-0.14	3.77	<0.001
Lagomorpha	33	OLS	(0)	12.9	9.8-17.1	17.84	<0.001	0.14	0.10-0.18	6.81	<0.001
		PGLS	1	19.2	9.6-38.6	8.29	<0.001	0.07	-0.03-0.17	1.38	0.178
			0.81	14.9	8.2-27.0	8.87	<0.001	0.11	0.02-0.20	2.45	0.020
Perissodactyla	16	OLS	(0)	86.2	28.3-262.8	7.84	<0.001	0.12	0.03-0.20	2.65	0.019
		PGLS	1	72.5	9.7-541.0	4.18	0.001	0.13	-0.03-0.28	1.64	0.122
			0.07	88.2	27.1-287.0	7.44	<0.001	0.11	0.02-0.21	2.47	0.027
Primates	140	OLS	(0)	66.5	56.9-77.8	52.63	<0.001	0.11	0.09-0.13	11.24	<0.001
		PGLS	1	84.4	59.7-119.4	25.08	<0.001	0.09	0.05-0.12	4.62	<0.001
			0.97*	81.2	59.5-110.9	27.67	<0.001	0.09	0.06-0.13	5.29	<0.001
Rodentia	385	OLS	(0)	12.2	10.8-13.7	40.70	<0.001	0.21	0.19-0.24	17.32	<0.001
		PGLS	1	29.1	22.3-37.8	25.07	<0.001	0.07	0.05-0.09	6.35	<0.001
			1.00*	29.1	22.3-37.8	25.07	<0.001	0.07	0.05-0.09	6.35	<0.001
Bovidae	106	OLS	(0)	63.7	47.6-85.3	27.89	<0.001	0.11	0.08-0.13	7.97	<0.001
		PGLS	1	93.5	66.1-132.2	25.67	<0.001	0.08	0.05-0.11	5.29	<0.001
			1.00*	93.5	66.1-132.2	25.67	<0.001	0.08	0.05-0.11	5.29	<0.001
Muridae	216	OLS	(0)	20.7	18.3-23.4	48.22	<0.001	0.06	0.03-0.09	3.70	<0.001
		PGLS	1	21.7	18.2-25.9	34.18	<0.001	0.04	0.01-0.07	2.64	<0.001
			1.00*	21.7	18.2-25.9	34.18	<0.001	0.04	0.01-0.07	2.64	<0.001
Sciuridae	68	OLS	(0)	33.6	25.6-44.2	25.29	<0.001	-0.01	-0.05-0.04	-0.23	0.815
		PGLS	1	30.1	22.3-40.7	22.11	<0.001	0.05	0.01-0.09	2.55	0.013
			0.98*	30.1	22.6-40.1	23.15	<0.001	0.05	0.01-0.09	2.55	0.013

* λ significantly different from 0

** λ significantly different from 0 and from 1

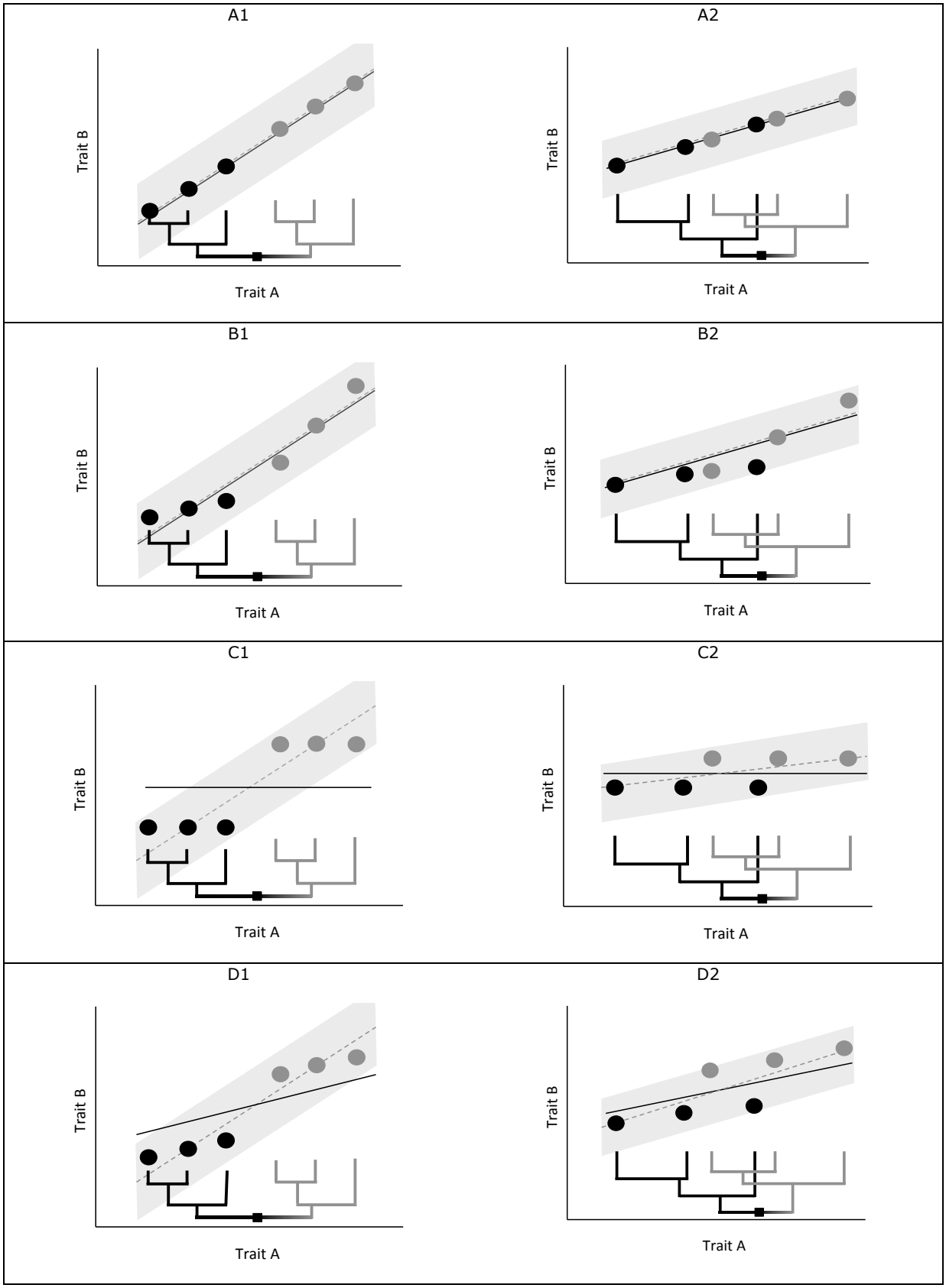
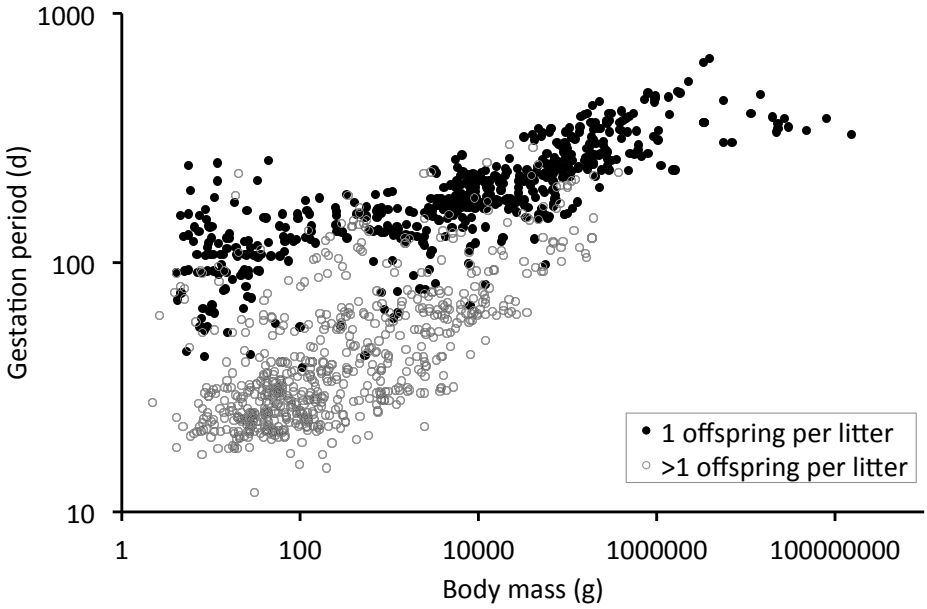


Figure 1. Potential combinations of results when applying statistical analyses without (result represented by the interrupted regression line) and with (result represented by the dark regression line) considering the phylogenetic structure of the sample (Harvey and Pagel 1991; Baker 2002; Nunn 2011). Dots represent data points for species that are linked by a hypothetical phylogeny with a basal node represented by the black square. Schemes on the left with dataset where different taxonomic levels are clearly separated in terms of their ranges for traits A and

B; schemes on the right allow for overlap of these ranges within taxonomic groups. A) both analyses yield the same (significant) result with similar scaling relationships; these relationships are identical at all taxonomic levels. This result strongly supports a universal scaling rule. B) both analyses yield the same (significant) result with similar scaling relationships, and the result is therefore indistinguishable from that of A; however, the result occurs because of a (apparently random) deviation of the scaling pattern within taxonomic subsets. This may support a universal scaling rule that allows for some 'deviation'. C) the analysis without accounting for phylogeny yields a significant result whereas the one that accounts for it does not, because there is no scaling pattern across closely related species. This result cannot be reconciled with the hypothesis that the scaling pattern is a physiological/ecological universal, because it differs between closely related species and at the level of the basal node. D) both analyses yield significant results, but the scaling relationships differ in their slope, because the scaling at the basal node is different from the scaling within closely related taxa. This result can also not be reconciled with the hypothesis that the scaling pattern is a physiological/ecological universal, but the scaling exponent may well be among closely related species. Note that in cases C and D, not only the slope of the relationship differs, but also the intercept; the intercept will have a larger confidence interval when accounting for phylogeny.



583

584 Figure 2. Body mass and gestation period length in eutherian mammals with one or more than one offspring per
585 litter.

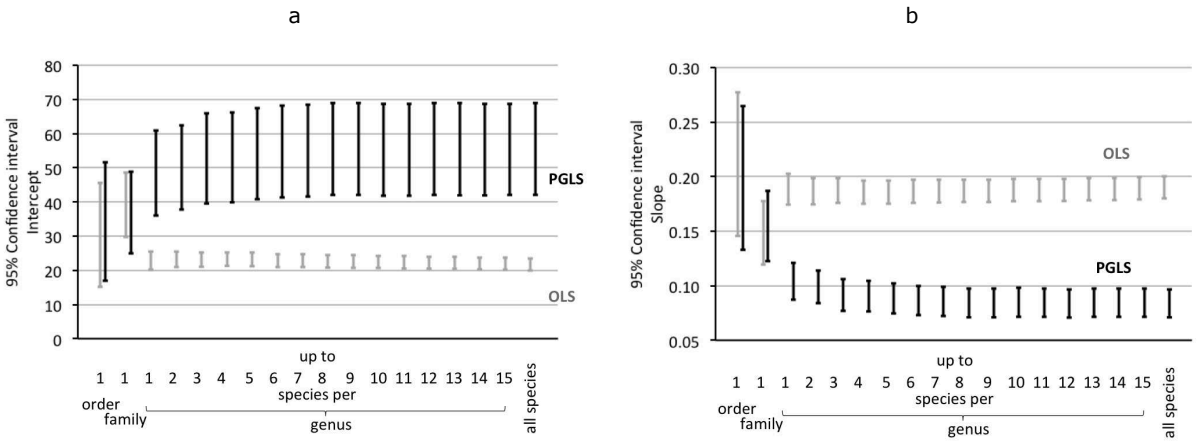


Figure 3. 95% confidence intervals of a) the intercept of the allometric relationship and b) the scaling exponent between gestation period (G) with body mass (M) according to $G = aM^b$ as analysed with statistical procedures that do not (OLS) and do (PGLS) account for the phylogenetic structure of the dataset in random data subsets allowing only a defined number of species per taxonomic level. n increased from 20 to 105, 574, 805, 927, 1006, 1054, 1092, 1124, 1148, 1165, 1177, 1185, 1193, 1197, 1200, 1203 and 1214. For the analyses at the order and family level, 100 randomized subsets were analysed each; for the analyses at the genus level, only one subset per species number step was analysed.